Comparing pharmacological treatments for cocaine dependence: Incorporation of methods for enhancing generalizability in meta-analytic studies

Ryoko Susukida\textsuperscript{1,2} | Rosa M. Crum\textsuperscript{1,3,4} | Hwanhee Hong\textsuperscript{1} | Elizabeth A. Stuart\textsuperscript{1,5,6} | Ramin Mojtabai\textsuperscript{1,4}

\textsuperscript{1} Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
\textsuperscript{2} Department of Mental Health Policy and Evaluation, National Institute of Mental Health, National Center of Neurology and Psychiatry, Kodaira, Tokyo, Japan
\textsuperscript{3} Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
\textsuperscript{4} Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
\textsuperscript{5} Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
\textsuperscript{6} Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

Correspondence
Ryoko Susukida, Department of Mental Health Policy and Evaluation, National Institute of Mental Health, National Center of Neurology and Psychiatry, 4-1-1 Ogawa-Higashi, Kodaira, Tokyo 187-8553, Japan.
Email: rsusukida@ncnp.go.jp

Funding information
National Institute on Drug Abuse (NIDA), Grant/Award Number: R01 DA036520

Abstract

Objectives: Few head-to-head comparisons of cocaine dependence medications exist, and combining data from different randomized controlled trials (RCTs) is fraught with methodological challenges including limited generalizability of the RCT findings. This study applied a novel meta-analytic approach to data of cocaine dependence medications.

Methods: Data from 4 placebo-controlled RCTs (Reserpine, Modafinil, Buspirone, and Ondansetron) were obtained from the National Institute of Drug Abuse Clinical Trials Network (\(n = 456\)). The RCT samples were weighted to resemble treatment-seeking patients (Treatment Episodes Data Set-Admissions) and individuals with cocaine dependence in general population (National Survey on Drug Use and Health). We synthesized the generalized outcomes with pairwise meta-analysis using individual-level data and compared the generalized outcomes across the 4 RCTs with network meta-analysis using study-level data.

Results: Weighting the data by the National Survey on Drug Use and Health generalizability weight made the overall population effect on retention significantly larger than the RCT sample effect. However, there was no significant difference between the population effect and the RCT sample effect on abstinence. Weighting changed the ranking of the effectiveness across treatments.

Conclusions: Applying generalizability weights to meta-analytic studies is feasible and potentially provides a useful tool in assessing comparative effectiveness of treatments for substance use disorders in target populations.

KEYWORDS
cocaine dependence treatment, generalizability, meta-analysis, network meta-analysis, propensity score weighting

1 INTRODUCTION

In 2014, the United States had approximately 913,000 individuals (0.29\% of the U.S. population) who met the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria for cocaine use disorder during the past 12 months (Center for Behavioral Health Statistics and Quality, 2015). According to the 2011 Drug Abuse Warning Network report (Substance Abuse and Mental Health Services Administration, 2013), nearly 40\% of drug misuse or abuse-related emergency room visits (505,224 out of 1.3 million visits) involved cocaine. Chronic cocaine use is associated with a number of adverse outcomes including psychotic symptoms (Morton, 1999), cardiovascular complications (Lange & Hillis, 2001), intracerebral hemorrhage (Buttner, 2012), and movement disorders such as Parkinson’s disease (Riezzo et al., 2012). It is also known that those with regular lifetime use of cocaine have significantly higher likelihood of premature deaths as compared with non-cocaine using peers (Qureshi, Chaudhry, & Suri, 2014).
There is clearly a need for evidence-based interventions for cocaine use disorders. There are a number of behavioral interventions shown to be effective for treating cocaine use disorders including contingency management (Kampman, 2010), cognitive-behavioral therapy (Penberthy, Ait-Daoud, Vaughan, & Fanning, 2010), and therapeutic communities (Vanderwert, Marshall, Nelson, Zeanah, & Fox, 2010). However, there are no pharmacological treatments for cocaine use disorder currently approved by the U.S. Food and Drug Administration (National Institute on Drug Abuse, 2016). A number of potential medications for cocaine use disorders have been examined in randomized clinical trials (RCTs; Castells et al., 2007; de Lima, de Oliveira Soares, Reisser, & Farrell, 2002; Kampman, 2005; Kampman, 2010; Penberthy et al., 2010). Existing studies targeted several neurobiological agents with putative effects on receptors considered to be involved in cocaine use disorder, such as dopamine, serotonin, gamma-aminobutyric acid, glutamate, and norepinephrine (Shorter, Domingo, & Kosten, 2015). The list of medications tried in past studies is long and includes the glutaminergic medication Modafinil (Dackis & O'Brien, 2003), gamma-aminobutyric acidergic medications such as baclofen, tiagabine, and topiramate (Kampman, 2005), disulfiram (Carroll et al., 2004), antidepressants such as desipramine (Arndt, McLellan, Dorozynsky, Woody, & O'Brien, 1994), and cocaine vaccination that produces antibodies against cocaine (Martell, Mitchell, Poling, Gonsal, & Kosten, 2005).

Despite some promising evidence of effectiveness in individual RCTs, previous pairwise meta-analyses of medications for cocaine use disorders have failed to produce evidence of overall treatment effectiveness of these medications or to identify clear advantages for one pharmacological agent (Castells et al., 2007; de Lima et al., 2002). These previous pairwise meta-analyses of medications for cocaine use disorders used a traditional approach of synthesizing study-level data typically obtained from publications. This approach makes it difficult to take into account the differences in the composition of RCT samples and to reliably compare different treatments.

Advances in meta-analytic methodology now make it possible to synthesize individual-level data from different RCTs (Riley, Lambert, & Abo-Zaid, 2010). Furthermore, the newly introduced method of network meta-analysis, also referred to as mixed treatment meta-analysis or multiple treatment comparison meta-analysis, now makes it possible to estimate comparative effectiveness across multiple interventions that are not evaluated against each other in any one study (Mills, Thorlund, & Ioannidis, 2013). The synthesis of multiple trials to assess comparative treatment effects is reasonable and transitive as long as different trials were conducted under similar clinical and methodological conditions where characteristics of the participants and the common comparator such as a placebo arm across trials could be assumed homogeneous (Cipriani, Higgins, Geddes, & Salanti, 2013).

Another major limitation of past RCTs for treatment of cocaine use disorder is the selective nature of the RCT sample (Sofuoglu, Dudish-Poulsen, Nicodemus, Babb, & Hatsuakami, 2000) that limits the external validity, or generalizability of the findings of the RCTs to the target population of individuals with cocaine use disorder in the general population, or to individuals in usual treatment settings. RCT findings are not directly generalizable to the target populations when there are treatment effect modifiers, and the distributions of the effect modifiers are different between the RCT samples and the target populations. This concern is not limited to cocaine treatment RCTs as there is a growing concern that the findings from RCTs in a number of health fields may not be generalizable to real world settings (Blanco et al., 2008a; Hoertel, Le Strat, Blanco, Lavaud, & Dubertret, 2012; Hoertel, Le Strat, Lavaud, Dubertret, & Limosin, 2013; Hoertel, Santiago, Wang, González-pinto, & Blanco, 2015; Okuda et al., 2010; Rothwell, 2005). However, the concerns may be amplified with regard to RCTs for treatments of substance use disorders (SUD) because of the stigma associated with such disorders and the specialized setting where treatments are offered may be an added challenge.

There is a growing body of research showing that individuals participating in RCTs are substantially different from the target populations (Blanco et al., 2008b; Humphreys, Weingardt, & Harris, 2007; Humphreys & Weisner, 2000; Okuda et al., 2010). According to a recent review (Moberg & Humphreys, 2016) that synthesized 15 studies examining the impact of SUD trial exclusion criteria on distributions of participants’ characteristics, commonly used exclusion criteria in SUD trials would exclude between 64% and 95% of potential participants. Furthermore, Susukida, Crum, Stuart, Ebnesajjad, and Mojtabai (2016) found substantial differences in distributions of characteristics between RCT samples and the target populations by comparing the characteristics of participants in 10 RCTs from the National Institute of Drug Abuse Clinical Trials Network (CTN) and the intended target populations consisting of people who seek treatment for SUD in usual care settings. A more recent study (Susukida, Crum, Ebnesajjad, Stuart, & Mojtabai, 2017) found that the significance of estimated sample treatment effects was different from that of the population effects when the distribution of characteristics of RCT samples was made to resemble the distribution of the target populations by using statistical weighting techniques. Most commonly, positive effects of trials in unweighted RCTs became statistically nonsignificant after weighting. To the best of our knowledge, however, no past studies of SUD treatments have synthesized data from individual RCTs with a view to the generalizability of results for the target population or have attempted to improve generalizability using statistical adjustments.

In this study, we embarked on a pairwise meta-analysis of individual-level data from four RCTs of medications to treat cocaine dependence as well as network meta-analysis of study-level data to compare the effects of these four treatments while incorporating generalizability of the findings of the RCTs to the target populations and adjusting the results to make them more generalizable. Two target populations were selected to investigate and enhance generalizability of the findings from meta-analyses: individuals seeking treatment for cocaine dependence at usual care settings and individuals with cocaine dependence in the general population, regardless of their treatment seeking behavior. Generalizing to these two diverse target samples addresses two distinct policy questions: (a) the efficacy of the treatment for individuals who seek treatment in usual care settings and (b) the efficacy of treatment if treatment is disseminated to the much wider population group who are not currently seeking any treatment but could potentially benefit from it.
2 | METHODS

2.1 | Data sources

RCT data were drawn from the National Institute on Drug Abuse Clinical Trials Network (2015) Data share Website. The National Institute of Drug Abuse CTN studies are nationwide multisite clinical trial studies to assess the effectiveness of SUD treatments. At the time of this writing, four data sets of RCTs of cocaine dependence medications were available (CTO0001, MDS0004, CTN00052, and CTO0005). CTO0001 (n = 119) examined the effectiveness of Reserpine, a dopamine depletory medication (T. Winhusen et al., 2008). MDS0004 (n = 210) examined the effectiveness of Modafinil, a non-amphetamine psychostimulant (Anderson et al., 2009). CTN00052 (n = 62) examined the effectiveness of Buspirone, an anxiolytic drug (T. M. Winhusen et al., 2014). CTO0005 (n = 65) examined the effectiveness of Ondansetron, a medicine mainly used for prevention of nausea (Johnson et al., 2006). We assumed that the synthesis of these trials was transitive because all four RCTs included a placebo arm and involved adults who met the DSM-Fourth Edition (American Psychiatric Association, 2000) for cocaine dependence. This study included a total of 456 patients from the four RCTs.

We selected two different target populations for generalizability weighting of the RCT samples. The first target population was drawn from the Treatment Episodes Data Set-Admissions (TEDS-A) in 2012 (the most recent wave of TEDS-A data available at the time of this writing). The TEDS-A is a part of the Behavioral Health Services Information System, which is maintained by the Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. More than 1.5 million individuals aged 12 years old or older who are admitted to public SUD treatment facilities are included in the TEDS-A every year. We limited the TEDS-A sample to patients with DSM-IV cocaine dependence who were 18 years old or older (n = 36,997) for generalizability to the treatment-seeking population.

The second target population was drawn from the National Survey on Drug Use and Health (NSDUH) 2013 and 2014 (the most recent waves of the survey available at the time of this writing). The NSDUH is an annual cross-sectional national survey also administered by the Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. Every year, NSDUH interviews a nationally representative sample of household residents 12 years old or older about their patterns of substance use and mental health problems. We limited the sample to those who met DSM-IV cocaine dependence criteria in the past year (n = 235) for generalizability to the cocaine-dependent individuals in the general population. The sampling weight was taken into account in every statistical analysis in this study because some demographic groups were oversampled in the NSDUH surveys (e.g., young adults).

2.2 | Measures

We identified eight comparable variables between the RCTs and the target populations with which to compute statistical weights for generalizing RCT outcomes to the two target populations: sex, race-ethnicity, age, educational attainment, employment status, marital status, intravenous drug use, and the number of past treatments for SUD.

We generalized the following two outcomes from the RCTs to the target populations: successful retention in the RCT and days of abstinence in the past 30 days. Successful retention in the RCT was defined as participation in the study until the end of the trial. Days of abstinence were calculated as the numbers of days without self-reported cocaine use in the prior 30 days.

We used the same outcomes across the four RCTs to allow us to compare how statistical weighting impacts the outcomes across the studies. The original investigators of the four RCTs reported on different outcomes. For example, whereas the original investigators of CTN00052 (T. M. Winhusen et al., 2014; the Buspirone trial) used maximum days of continuous cocaine abstinence as the primary outcome, the investigators of CTO0005 (Johnson et al., 2006; the Ondansetron trial) used percentage of study participants with a cocaine-free week as their primary outcome. The observed estimates of RCT findings in this study were not necessarily the same as the findings in published reports by original investigators (Anderson et al., 2009; Johnson et al., 2006; T. Winhusen et al., 2007; T. M. Winhusen et al., 2014).

2.3 | Statistical analysis

Analyses were conducted in four stages. First, we compared the eight characteristics noted above between the RCT samples and the target populations. Pearson’s chi-squared tests were conducted to examine if there were significant differences in the distributions of the eight variables between the RCT samples and the target populations. As the TEDS-A includes a significant amount of missing data (12.7%), we used multiple imputation with the STATA ice command (version 13; Royston, 2004) and created 50 imputed data sets. The percentage of missing observations in each RCT and target populations is presented in Table A1. There were no missing data in the NSDUH sample.

Second, to generalize the results from the RCTs to the target populations, we used a weighting-based method, which weights RCT samples to resemble the target populations (Cole & Stuart, 2010; Stuart, Bradshaw, & Leaf, 2015; Stuart, Cole, Bradshaw, & Leaf, 2011). This approach is similar to the inverse probability weighting method, which is often used for nonexperimental studies (Mansournia & Altman, 2016). We first estimated propensity score of RCT participation, p, which was defined as the conditional probability of an individual participating in the RCTs based on the eight variables described above. The mean propensity score across 50 imputed data sets was used for TEDS-A to take into account the missing data. To calculate the propensity scores, we used a nonparametric random forests approach, using the “randomForest” (Liaw & Wiener, 2002) package in R (R Development Core Team, 2016). Although the random forests approach has some advantages over a parametric approach such as a higher predictive accuracy and the ability to reduce extreme propensity scores (Strobl, Malley, & Tutz, 2009), we still encountered some outlying values for propensity scores and trial participation weights. In order to improve the performance of the propensity score-based weighting, we used weight trimming, also referred as truncation, in which we replaced extreme large values at the 95th percentile values following
the method introduced in a study by Lee, Lessler, and Stuart (2011). To assess generalizability of the findings of each RCT, we computed trial participation weights for each trial as \((1 - p)/p\). We conducted weighted regression analyses with the weights for each trial, \((1 - p)/p\), using the STATA pweights command (version 13).

Third, we conducted unweighted and weighted pairwise meta-analyses of the four RCTs to estimate the overall treatment effect of cocaine dependence medications, which is a pooled treatment effect of four medications assessed through separately conducted RCTs as compared with placebo arms. The unweighted analyses estimated the sample treatment effects whereas the weighted analyses estimated population-generalized effects. For the binary outcome of retention, we estimated risk differences for treatment effects; whereas, for the continuous measure of days of abstinence in the past 30 days, we used linear regression models. Baseline variables in the trial samples were not adjusted, assuming that randomization was successful in each trial. We estimated the pooled treatment effect with clustered standard errors at each trial level. In addition to comparing the statistical significance of the treatment effects from unweighted and weighted regression models, we also statistically compared the treatment effect sizes of unweighted and weighted models, using the STATA suet (seemingly unrelated estimation) command (Weesie, 1999). Furthermore, to explore potential reasons for differences in sample and generalized treatment effects, we conducted a series of subgroup analyses in which we estimated treatment effects in subgroups that were overrepresented or underrepresented in the RCT samples compared with the target populations.

Fourth, to directly compare the effects of the medications, we conducted an unweighted and weighted network meta-analysis of the four RCTs to estimate the comparative treatment effects across the four medications. Past research has shown that statistical precision of estimated treatment effects from network meta-analyses is often better than that of estimated effects from pairwise comparisons in meta-analysis when the distributions of effect modifiers are balanced (Cipriani et al., 2013; Jansen & Naci, 2013; Salanti, 2012). Network meta-analysis also allows for determination of relative rankings of multiple treatments (Salanti, Ades, & Ioannidis, 2011). We fitted fixed-effect network meta-analysis models with the “gemtc” package in R (Dias, Welton, Caldwell, & Ades, 2010).

## RESULTS

### 3.1 Comparison of characteristics of four CTN trials and target populations

Table 1 presents the comparison of characteristics between the four RCTs and the TEDS-A target population. Overall, the RCT samples had significantly lower proportions of women, non-Hispanic White individuals, and patients younger than 35 years old than the TEDS-A; whereas, the RCT samples had significantly higher proportions of individuals with 12 or more years of education, individuals with full-time jobs, married individuals, and individuals who used intravenous drugs than the TEDS-A. For all the RCTs, the proportions of patients with full-time jobs were significantly higher than the TEDS-A.

Table 2 presents the comparison of characteristics between the four RCTs and the NSDUH target population. Overall, the RCT samples had significantly lower proportions of women, individuals

### Table 1

Comparison of baseline characteristics (%) of four Clinical Trial Network (CTN) trials and the Treatment Episode Data Set-Admissions (TEDS-A)

<table>
<thead>
<tr>
<th></th>
<th>Target population TEDS-A, 2012 N = 36,997</th>
<th>Randomized controlled trials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall N = 456</td>
<td>CTO0001 (1) N = 119</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>42.3%</td>
<td>27.9%</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>41.2%</td>
<td>27.4%</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;35</td>
<td>33.8%</td>
<td>19.1%</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥12 years</td>
<td>66.4%</td>
<td>82.8%</td>
</tr>
<tr>
<td><strong>Employment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time</td>
<td>9.2%</td>
<td>33.8%</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>12.3%</td>
<td>23.3%</td>
</tr>
<tr>
<td><strong>IV drug use</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>5.7%</td>
<td>9.4%</td>
</tr>
<tr>
<td><strong>Prior treatments</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>64.3%</td>
<td>61.0%</td>
</tr>
</tbody>
</table>

Note. Pearson’s \(\chi^2\) test was conducted. Numbers shown in bold type indicate statistically significant differences between randomized controlled trial and TEDS-A samples at \(p < .05\).
from non-Hispanic White racial-ethnic groups, and those younger than 35 years old than the NSDUH; whereas, the trial samples had significantly higher proportions of married individuals than the NSDUH.

3.2 | Pairwise meta-analysis

Table 3 presents the results of the pairwise meta-analyses for the overall treatment effect on trial retention and abstinence. Risk ratios and mean differences ($\beta$s) from both unweighted and weighted regression models, respectively, are presented with the 95% confidence intervals. In addition, the results of comparisons of regression coefficients are presented in Table 3.

For retention, the overall population treatment effect was significantly larger for the analyses weighted by the NSDUH target population than the sample treatment effect (Table 3). A risk ratio of 1.41 suggests that individuals receiving the active pharmaceutical agents have 1.41 times the likelihood of being retained in the study compared with those treated with placebo. In contrast, there was no statistically significant difference between the population-adjusted effect and the sample effect on self-reported abstinence (Table 3).

### TABLE 2 Comparison of baseline characteristics (%) of four Clinical Trial Network (CTN) trials and the National Survey on Drug Use and Health (NSDUH)

<table>
<thead>
<tr>
<th>Target population</th>
<th>Randomized controlled trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSDUH, 2013-2014</td>
<td>Overall</td>
</tr>
<tr>
<td>N=235</td>
<td>N=456</td>
</tr>
<tr>
<td><strong>Total number</strong></td>
<td>235</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td>37.4</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td>52.8</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>49.9</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td>79.4</td>
</tr>
<tr>
<td><strong>Employment</strong></td>
<td>33.7</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td>15.5</td>
</tr>
<tr>
<td><strong>IV drug use</strong></td>
<td>6.3</td>
</tr>
<tr>
<td><strong>Prior treatments</strong></td>
<td>65.5</td>
</tr>
</tbody>
</table>

Note. Pearson’s $\chi^2$ test was conducted. Numbers shown in bold type indicate statistically significant differences between randomized controlled trial and NSDUH samples at $p < .05$. The sampling weight was taken into consideration for the NSDUH target population.

### TABLE 3 Unweighted and weighted meta-analysis of pharmacological cocaine dependence treatments with two target populations

| Target population = the Treatment Episode Data Set-Admissions (TEDS-A) |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| **Outcome** | **Unweighted RR** | **95% CI** | **Weighted RR** | **95% CI** |
| Retention | 1.08* | 1.00, 1.16 | 1.19** | 1.08, 1.30 |
| **Outcome** | **Unweighted $\beta$** | **95% CI** | **Weighted $\beta$** | **95% CI** |
| Abstinence | -0.80 | -3.02, 1.42 | -1.12 | -2.28, 0.05 |

| Target population = the National Survey on Drug Use and Health (NSDUH) |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| **Outcome** | **Unweighted RR** | **95% CI** | **Weighted RR** | **95% CI** |
| Retention | 1.08* | 1.00, 1.16 | 1.41** | 1.11, 1.78 |
| **Outcome** | **Unweighted $\beta$** | **95% CI** | **Weighted $\beta$** | **95% CI** |
| Abstinence | -0.80 | -3.02, 1.42 | -0.84 | -3.12, 1.44 |

Note. RR stands for risk ratio. Weighted results were weighted by the TEDS-A and NSDUH target populations. The standard errors were clustered at the trial level. The weight was truncated at 95 percentiles to eliminate extreme weights. CI = confidence intervals.

**p < .01.
*p < .05.
The results of subgroup analysis of treatment effects are presented in Table A2. We found some consistent patterns in the directions of change in outcomes through weighting and by subgroup analysis. For example, the non-married individuals that were slightly underrepresented in the overall RCTs (p = .10) showed evidence of larger treatment effects on retention than the overrepresented group of married individuals. Weighting the RCT samples to resemble the NSDUH target population increased the weights for the subsample of non-married individuals that had larger treatment effect sizes, leading to a significantly larger treatment effect on retention after weighting with the NSDUH target population.

### 3.3 Network meta-analysis

Table 4 presents the results of network meta-analysis comparing the effect of the four medications on cocaine dependence. Risk ratios and mean differences (βs) from both unweighted and weighted regression models, respectively, are presented with the 95% confidence intervals. We also present the relative rankings of the treatments, computed as the probabilities of each treatment being the best among all the treatments in the network meta-analysis.

In the unweighted model for each medication, there was no significant treatment effect. Although it should be noted that the 95% confidence intervals for the estimated treatment effects for each medication overlapped with each other, Ondansetron was the most efficacious medication for retention, and Modafinil was the most efficacious medication for abstinence. Weighting altered the relative ranking of the treatments. For retention, weighting by TEDS sample showed evidence of larger treatment effects on abstinence than underrepresented group of married individuals. Weighting the RCT samples to resemble the NSDUH target population increased the weights for the subsample of non-married individuals with larger treatment effect sizes, leading to a smaller treatment effect on abstinence.

### 4 DISCUSSIONS

This study showed that the findings from meta-analytic studies of cocaine dependence medications may not be directly applicable to potential target populations. The estimated overall target population-weighted treatment effect of four cocaine dependence medications on retention was significantly larger than the treatment effect from the RCTs when generalized by the NSDUH data, representing the target population of individuals with cocaine dependence in the community. Weighting the RCT samples to resemble target populations also altered the relative ranking of the efficacy across different medications. The results from the subgroup analysis of treatment effects partially explained these differences in effect estimates between unweighted and weighted meta-analyses.

Stuart et al. (2015) used the same weighting-based approach to generalize the results of a school-based behavioral intervention trial and found that the estimated population effect of the intervention was slightly attenuated compared with the estimated sample effect from the RCT. In the context of SUD RCTs, Susukida et al. (2017) demonstrated a statistically significant difference between the sample

<table>
<thead>
<tr>
<th>TABLE 4 Unweighted and weighted network meta-analysis of four pharmacological treatments for cocaine dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retention</td>
</tr>
<tr>
<td>RR</td>
</tr>
<tr>
<td>Reserpine vs. placebo</td>
</tr>
<tr>
<td>Modafinil vs. placebo</td>
</tr>
<tr>
<td>Buspirone vs. placebo</td>
</tr>
<tr>
<td>Ondansetron vs. placebo</td>
</tr>
</tbody>
</table>

| Abstinence | Unweighted | Weighted (with TEDS-A) | Weighted (with NSDUH) |
|---------------------------------------------------------------|
| RR  | 95% CI | Rank (%) | RR  | 95% CI | Rank (%) | RR  | 95% CI | Rank (%) |
| Reserpine vs. placebo | −2.63 | −6.21, 0.94 | 2.3 | −2.13 | −7.42, 3.16 | 15.0 | −1.44 | −8.22, 5.32 | 23.0 |
| Modafinil vs. placebo | 1.14 | −1.63, 3.90 | 61.9 | −0.09 | −3.95, 3.78 | 40.0 | 0.74 | −3.43, 4.92 | 50.1 |
| Buspirone vs. placebo | −2.15 | −5.50, 1.21 | 3.2 | −2.62 | −5.57, 0.32 | 2.0 | −3.65 | −8.60, 1.29 | 3.1 |
| Ondansetron vs. placebo | −1.82 | −9.12, 5.46 | 20.6 | −5.05 | −13.41, 3.24 | 8.5 | −4.07 | −9.26, 1.07 | 2.6 |

Note. RR stands for risk ratio. Weighted results were weighted by the Treatment Episode Data Set-Admissions (TEDS-A) and National Survey on Drug Use and Health (NSDUH) target populations. The weight was truncated at 95 percentiles to eliminate extreme weights. Rank (%) represents the estimated probability of each medication being the most efficacious medication among four medications. CI = confidence intervals.

*p < .05.
The findings from this study have implications for future meta-analytic studies of SUD treatments. As shown in this study, the overall treatment effect size and comparative effects changed when the deviations of each RCT sample from the target population were taken into account. Unlike the previous study by Susukida et al. (2017) that found a decrease in treatment effect sizes after weighting of data from 10 SUD RCT samples, the effect size associated with the cocaine dependence medications on retention in the present study became significantly larger after weighting by the NSDUH population weight. This implies that the effect of weighting-based methods may vary depending on how and to what extent the composition of the RCT samples and target populations vary. Furthermore, differences in efficacies among different treatments for the same condition may be impacted by the compositions of the RCT samples for each treatment. Target population weighting of the RCTs changed the relative ranking of treatments for cocaine dependence in this study although it should be noted that all the 95% confidence intervals for each treatment effect estimate from each RCT overlapped, indicating low confidence regarding differences in treatment effects across the four RCTs. This study also provided some evidence that suggests that the mechanisms through which the population treatment effects were different from the sample effects could be partially explained by treatment effect heterogeneity among underrepresented or overrepresented subgroups of individuals in the RCTs.

Several limitations should be taken into account when interpreting the findings of this study. First, the number of characteristics recorded in both the clinical trial samples and the target populations was relatively small, which likely limited our ability to account for potentially important treatment effect modifiers that may have been different between the RCTs and the target populations, such as motivation for treatment, severity of SUD, and comorbidity. Future studies would benefit from a method recently introduced by Nguyen, Ebnesajjad, Cole, and Elizabeth (2017), which allows researchers to estimate the population treatment effects in cases where treatment effect modifiers are not observed in target populations, but only in RCTs. Second, the weighting-based approach might not have made the distributions of the clinical trial samples sufficiently close to the distributions of the target populations to estimate the population treatment effects because of the substantial differences between the clinical trial samples and the corresponding target populations. Weighting-based approach in this study was not able to fully balance the distributions of the RCT samples and the target populations. As shown in Table A3, there were statistically significant differences in the estimated propensity scores between each RCT and its target populations. The weighting-based method is more suitable to estimate the population treatment effect when the distributions of characteristics in RCTs overlap with those of the target populations. Third, the present study could only show suggestive evidence of treatment effect heterogeneity across subgroups of individuals in the clinical trial samples because the CTN studies did not originally intend to assess the treatment effect heterogeneity, and the subgroup analyses conducted here were not sufficiently powered. Lastly, the number of trials included in this study was limited, which likely limited the reliability of the network meta-analysis (Jansen et al., 2011) and our ability to conclude which medication is the most promising for treating cocaine dependence.

Limitations notwithstanding, the findings of this study provide insight into the generalizability of meta-analysis of cocaine dependence medications. The overall population treatment effect on trial retention appears promising for the four cocaine dependence medications (Reserpine, Modafinil, Buspirone, and Ondansetron). The relative ranking of effectiveness among the four treatments was altered when we considered generalizability of the findings to the target populations. With the growing number of RCTs for cocaine dependence medications, future meta-analytic studies should assess overall treatment effects and comparative effectiveness while considering generalizability to target populations. The weighting-based approach used in this study is applicable to meta-analyses of clinical trials of other SUD treatments, as well as other health interventions, especially when generalizability of the findings is a concern. Although an increasing number of clinical trials for SUD treatments use less stringent inclusion and exclusion criteria, reducing concerns about generalizability of findings (Wang, Ulbricht, & Schoenbaum, 2009), it may not always be possible to recruit samples that are fully representative of the target populations, for example, when there are safety concerns for certain population subgroups. In these circumstances, the weighting-based method used in this study could be useful to assess applicability of the findings to treatment seeking target populations or to all individuals with the specific condition in the general community.

DECLARATION OF INTEREST STATEMENT

Dr. Susukida and Dr. Hong have nothing to disclose. Drs. Crum, Stuart, and Mojtabai report grants from the National Institute on Drug Abuse, National Institute of Mental Health, and/or the National Institute on Alcohol Abuse and Alcoholism during the conduct of the study. Dr. Mojtabai has received research funding and consulting fees from Bristol-Myers Squibb and Lundbeck Pharmaceuticals.

ACKNOWLEDGEMENTS

This work was supported by Grant R01 DA036520 from the National Institute on Drug Abuse (NIDA). The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

ORCID

Ryoko Susukida https://doi.org/10.1002/jps.3080051129

REFERENCES


How to cite this article: Susukida R, Crum RM, Hong H, Stuart EA, Mojtabai R. Comparing pharmacological treatments for cocaine dependence: Incorporation of methods for enhancing generalizability in meta-analytic studies. Int J Methods Psychiatr Res. 2018;e1609. https://doi.org/10.1002/mpr.1609
# APPENDIX A

## TABLE A1
Number and percentage of cases with missing values for each covariate in the four cocaine dependence clinical trials and the target populations

<table>
<thead>
<tr>
<th>Covariate</th>
<th>TEDS-A, 2012 Target</th>
<th>NSDUH, 2013-2014 Target</th>
<th>CTO0001 RCT</th>
<th>MDS0004 RCT</th>
<th>CTN0052 RCT</th>
<th>CTO0005 RCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total valid N</td>
<td>32,287</td>
<td>235</td>
<td>115</td>
<td>209</td>
<td>62</td>
<td>65</td>
</tr>
<tr>
<td>Total missing N</td>
<td>4,710</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% total missing</td>
<td>12.7</td>
<td>0.0</td>
<td>3.3</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid N</td>
<td>36,919</td>
<td>235</td>
<td>119</td>
<td>210</td>
<td>62</td>
<td>65</td>
</tr>
<tr>
<td>Missing N</td>
<td>78</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% missing</td>
<td>0.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid N</td>
<td>36,734</td>
<td>235</td>
<td>117</td>
<td>209</td>
<td>62</td>
<td>65</td>
</tr>
<tr>
<td>Missing N</td>
<td>263</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% missing</td>
<td>0.7</td>
<td>0.0</td>
<td>1.7</td>
<td>0.5</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid N</td>
<td>36,997</td>
<td>235</td>
<td>119</td>
<td>210</td>
<td>62</td>
<td>65</td>
</tr>
<tr>
<td>Missing N</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% missing</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid N</td>
<td>36,281</td>
<td>235</td>
<td>119</td>
<td>210</td>
<td>62</td>
<td>65</td>
</tr>
<tr>
<td>Missing N</td>
<td>716</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% missing</td>
<td>0.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Employment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid N</td>
<td>35,965</td>
<td>235</td>
<td>119</td>
<td>210</td>
<td>62</td>
<td>65</td>
</tr>
<tr>
<td>Missing N</td>
<td>1,032</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% missing</td>
<td>2.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid N</td>
<td>34,428</td>
<td>235</td>
<td>118</td>
<td>210</td>
<td>62</td>
<td>65</td>
</tr>
<tr>
<td>Missing N</td>
<td>2,569</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% missing</td>
<td>6.9</td>
<td>0.0</td>
<td>0.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>IV drug use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid N</td>
<td>36,328</td>
<td>235</td>
<td>119</td>
<td>210</td>
<td>62</td>
<td>65</td>
</tr>
<tr>
<td>Missing N</td>
<td>669</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% missing</td>
<td>0.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Number of prior treatments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valid N</td>
<td>35,800</td>
<td>235</td>
<td>118</td>
<td>210</td>
<td>62</td>
<td>65</td>
</tr>
<tr>
<td>Missing N</td>
<td>1,197</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% missing</td>
<td>0.3</td>
<td>0.0</td>
<td>0.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Note. TEDS-A = Treatment Episode Data Set-Admissions; NSDUH = National Survey on Drug Use and Health; RCT = randomized controlled trial; CTN = Clinical Trials Network.
### TABLE A2  Results of subgroup analysis of treatment effects

<table>
<thead>
<tr>
<th>Study (treatment)</th>
<th>Variable</th>
<th>Retention (end of trial)</th>
<th>Abstinence (end of trial)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>Overall (four RCTs, N = 456)</td>
<td>Sex</td>
<td>Men</td>
<td>69.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Women</td>
<td>71.6</td>
</tr>
<tr>
<td></td>
<td>Race</td>
<td>White</td>
<td>65.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-White</td>
<td>71.4</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>&lt;35</td>
<td>59.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥35</td>
<td>72.4</td>
</tr>
<tr>
<td></td>
<td>Education</td>
<td>&lt;12 years</td>
<td>75.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥12 years</td>
<td>68.8</td>
</tr>
<tr>
<td></td>
<td>Employment</td>
<td>Full-time</td>
<td>58.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Other</td>
<td>76.1</td>
</tr>
<tr>
<td></td>
<td>Marital status</td>
<td>Married</td>
<td>56.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-married</td>
<td>73.7</td>
</tr>
<tr>
<td></td>
<td>IV drug use</td>
<td>Yes</td>
<td>76.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td>69.0</td>
</tr>
<tr>
<td></td>
<td>Prior treatment</td>
<td>Yes</td>
<td>76.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No</td>
<td>60.5</td>
</tr>
</tbody>
</table>

CTO0001 (Reserpine, N = 119)

| Sex | Men | 67.4 | 68.3 | χ² = 0.0 | — | 20.6 | 23.2 | t = 1.3 | — |
|     | Women | 64.7 | 55.6 | χ² = 0.3 | p = .61 | 16.8 | 19.1 | t = 0.6 | p = .92 |
| Race | White | 61.1 | 44.4 | χ² = 0.7 | — | 20.6 | 29.5 | t = 1.3 | — |
|      | Non-White | 69.3 | 69.4 | χ² = 0.0 | p = .47 | 18.9 | 21.8 | t = 1.4 | p = .36 |
| Age | <35 | 66.7 | 46.7 | χ² = 1.1 | — | 11.4 | 20.7 | t = 2.1 | — |
|      | ≥35 | 66.7 | 70.5 | χ² = 0.2 | p = .28 | 21.6 | 22.5 | t = 0.5 | p = .05 |
| Education | <12 years | 62.5 | 70.0 | χ² = 0.1 | — | 22.2 | 22.1 | t = 0.0 | — |
|      | ≥12 years | 67.3 | 63.3 | χ² = 0.2 | p = .64 | 19.2 | 22.2 | t = 1.5 | p = .56 |
| Employment | Full-time | 54.6 | 66.7 | χ² = 0.5 | — | 18.2 | 25.0 | t = 2.3 | — |
|      | Other | 73.7 | 63.6 | χ² = 1.0 | p = .25 | 20.1 | 21.2 | t = .5 | p = .15 |
| Marital status | Married | 46.2 | 70.0 | χ² = 1.3 | — | 17.5 | 22.3 | t = 1.1 | — |
|      | Non-married | 71.7 | 63.3 | χ² = 0.8 | p = .16 | 20.1 | 22.2 | t = 1.0 | p = .59 |
| IV drug use | Yes | 100.0 | 80.0 | χ² = 0.7 | — | 25.0 | 17.8 | t = −1.2 | — |
|      | No | 64.9 | 63.0 | χ² = 0.0 | p = .19 | 19.1 | 22.7 | t = 1.9 | p = .10 |
| Prior treatment | Yes | 67.7 | 67.6 | χ² = 0.0 | — | 19.5 | 23.3 | t = 1.7 | — |
|      | No | 65.5 | 57.1 | χ² = 0.4 | p = .67 | 19.6 | 20.2 | t = 0.2 | p = .39 |
### TABLE A2 (Continued)

<table>
<thead>
<tr>
<th></th>
<th>MDS0004 (Modafinil, N = 210)</th>
<th>CTN0052 (Buspirone, N = 62)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Retention (end of trial)</td>
<td>Abstinence (end of trial)</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>69.0</td>
<td>58.8</td>
</tr>
<tr>
<td>Women</td>
<td>76.3</td>
<td>71.4</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>73.0</td>
<td>62.5</td>
</tr>
<tr>
<td>Non-White</td>
<td>70.3</td>
<td>61.7</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;35</td>
<td>59.1</td>
<td>50.0</td>
</tr>
<tr>
<td>≥35</td>
<td>73.3</td>
<td>64.1</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;12 years</td>
<td>76.5</td>
<td>60.0</td>
</tr>
<tr>
<td>≥12 years</td>
<td>70.3</td>
<td>62.7</td>
</tr>
<tr>
<td><strong>Employment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time</td>
<td>62.2</td>
<td>47.8</td>
</tr>
<tr>
<td>Other</td>
<td>75.5</td>
<td>69.4</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>70.6</td>
<td>76.2</td>
</tr>
<tr>
<td>Non-married</td>
<td>71.2</td>
<td>56.9</td>
</tr>
<tr>
<td><strong>IV drug use</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>80.0</td>
<td>33.3</td>
</tr>
<tr>
<td>No</td>
<td>69.9</td>
<td>63.8</td>
</tr>
<tr>
<td><strong>Prior treatment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>79.3</td>
<td>63.4</td>
</tr>
<tr>
<td>No</td>
<td>58.9</td>
<td>61.3</td>
</tr>
<tr>
<td><strong>CTN0052</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>95.8</td>
<td>80.0</td>
</tr>
<tr>
<td>Women</td>
<td>81.8</td>
<td>91.7</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>88.9</td>
<td>71.4</td>
</tr>
<tr>
<td>Non-White</td>
<td>92.3</td>
<td>90.0</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;35</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td>≥35</td>
<td>90.6</td>
<td>88.5</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;12 years</td>
<td>92.9</td>
<td>100.0</td>
</tr>
<tr>
<td>≥12 years</td>
<td>90.5</td>
<td>79.0</td>
</tr>
<tr>
<td><strong>Employment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time</td>
<td>85.7</td>
<td>33.3</td>
</tr>
<tr>
<td>Other</td>
<td>92.2</td>
<td>100.0</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>66.7</td>
<td>60.0</td>
</tr>
<tr>
<td>Non-married</td>
<td>93.8</td>
<td>90.9</td>
</tr>
<tr>
<td><strong>IV drug use</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>No</td>
<td>90.9</td>
<td>84.6</td>
</tr>
<tr>
<td><strong>Prior treatment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>91.4</td>
<td>85.2</td>
</tr>
<tr>
<td>No</td>
<td>89.7</td>
<td>79.2</td>
</tr>
<tr>
<td></td>
<td>Retention (end of trial)</td>
<td>Abstinence (end of trial)</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td><strong>CTO0005 (Ondansetron, N = 65)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>56.1</td>
<td>50.0</td>
</tr>
<tr>
<td>Women</td>
<td>50.0</td>
<td>0.0</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>47.6</td>
<td>25.0</td>
</tr>
<tr>
<td>Non-White</td>
<td>60.7</td>
<td>50.0</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;35</td>
<td>58.6</td>
<td>50.0</td>
</tr>
<tr>
<td>≥35</td>
<td>50.0</td>
<td>33.3</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;12 years</td>
<td>55.0</td>
<td>46.2</td>
</tr>
<tr>
<td>≥12 years</td>
<td>55.6</td>
<td>33.3</td>
</tr>
<tr>
<td><strong>Employment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time</td>
<td>50.0</td>
<td>25.0</td>
</tr>
<tr>
<td>Other</td>
<td>61.9</td>
<td>62.5</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>28.6</td>
<td>50.0</td>
</tr>
<tr>
<td>Non-married</td>
<td>65.7</td>
<td>40.0</td>
</tr>
<tr>
<td><strong>IV drug use</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>–</td>
<td>64.3</td>
</tr>
<tr>
<td>No</td>
<td>51.3</td>
<td>43.8</td>
</tr>
<tr>
<td><strong>Prior treatment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>52.6</td>
<td>20.0</td>
</tr>
<tr>
<td>No</td>
<td>58.6</td>
<td>54.6</td>
</tr>
</tbody>
</table>

Note. RCTs = randomized controlled trials; CTN = Clinical Trials Network. Bold letters indicate that the test statistics were statistically significant at $p<0.05$.

*All the participants who had IV drug use in the treatment arm successfully retained in the study, and it was not possible to estimate a coefficient for interaction term.

*All the participants aged <35 years old in the treatment arm successfully retained in the study, and it was not possible to estimate a coefficient for interaction term.

*All the participants with less than 12 years of education in the control arm successfully retained in the study, and it was not possible to estimate a coefficient for interaction term.

*All the participants without full-time job in the control arm successfully retained in the study, and it was not possible to estimate a coefficient for interaction term.

*All the participants who had IV drug use successfully retained in the study, and it was not possible to calculate chi-squared statistics and to estimate a coefficient for interaction term.

*There was no observation of those who had no prior treatments.

*No female participants in the control arm successfully retained in the study, and it was not possible to estimate a coefficient for interaction term.

*All the participants who had IV drug use were in the treatment arm, and it was not possible to calculate chi-squared statistics and to estimate a coefficient for interaction term.

*There was only one observation in the control arm, and it was not possible to calculate $t$ statistics.

*There was no observation in the control arm, and it was not possible to calculate $t$ statistics and to estimate a coefficient for interaction term.
<table>
<thead>
<tr>
<th>Study number</th>
<th>Intervention</th>
<th>RCT</th>
<th>TEDS-A</th>
<th>Δp&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Pooled standard deviation</th>
<th>Standard Δp&lt;sup&gt;b&lt;/sup&gt;</th>
<th>t test</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTO0001</td>
<td>Reserpine</td>
<td>0.64</td>
<td>0.25</td>
<td>0.39</td>
<td>0.21</td>
<td>1.86</td>
<td>20.41</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>MD50004</td>
<td>Modafinil</td>
<td>0.60</td>
<td>0.24</td>
<td>0.36</td>
<td>0.22</td>
<td>1.61</td>
<td>23.44</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>CTN0052</td>
<td>Buspirone</td>
<td>0.66</td>
<td>0.24</td>
<td>0.42</td>
<td>0.22</td>
<td>1.94</td>
<td>15.30</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>CTO0005</td>
<td>Ondansetron</td>
<td>0.73</td>
<td>0.17</td>
<td>0.56</td>
<td>0.19</td>
<td>3.01</td>
<td>24.44</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study number</th>
<th>Intervention</th>
<th>RCT</th>
<th>NSDUH</th>
<th>Δp&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Pooled standard deviation</th>
<th>Standard Δp&lt;sup&gt;b&lt;/sup&gt;</th>
<th>t test</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTO0001</td>
<td>Reserpine</td>
<td>0.73</td>
<td>0.24</td>
<td>0.49</td>
<td>0.35</td>
<td>1.40</td>
<td>16.61</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>MD50004</td>
<td>Modafinil</td>
<td>0.74</td>
<td>0.31</td>
<td>0.43</td>
<td>0.36</td>
<td>1.22</td>
<td>16.25</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>CTN0052</td>
<td>Buspirone</td>
<td>0.80</td>
<td>0.23</td>
<td>0.57</td>
<td>0.34</td>
<td>1.68</td>
<td>16.05</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>CTO0005</td>
<td>Ondansetron</td>
<td>0.61</td>
<td>0.31</td>
<td>0.30</td>
<td>0.25</td>
<td>1.20</td>
<td>9.86</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Note: CTN = Clinical Trials Network; RCT = randomized control trial.

<sup>a</sup> Δp is difference between propensity scores of the RCT sample and the target population.

<sup>b</sup> Standardized Δp is computed as Δp divided by pooled standard deviation.